BLUPs and EBLUPs

Tom Struppeck

CLRS 2016

Why do we like Least Squares Regression?

The Gauss-Markov Theorem tells us that when we estimate the coefficients using OLS regression, the resulting estimates are BLUE.

- Best: Smallest variance within this class of estimators
- Linear: The estimates are linear functions of the data (the response)
- Unbiased: The expected value of the estimates is the true value
- Estimator: A function of the data and known constants.

There are some hypotheses for Gauss-Markov

- The explanatory variables are assumed to be known with no error.
- If there are errors in the explanatory variables, the slope coefficients may be biased towards zero.
- (And so would not be BLUE)

When we have a mixed model …

- We slightly change what the letters stand for:

- BLUP
- Best: Smallest MSE among estimators in this class
- Linear: Linear in the data (no change from BLUE)
- Unbiased: The expected value of the estimate is the expected value of the true value
- Predictor: Since the random effects are random variables, we don't "estimate" them; instead we "predict" the response variable

In order to compute the BLUP, we need:

- To know the variance matrix. In practice, we seldom know this, so we estimate it.
- If we estimate it from the same data that we are using to fit our model, we call the resulting estimate the EBLUP (empirical BLUP)
- This is analogous to the standard error vs the standard deviation

An old (1950) example

- Suppose that IQ scores are normally distributed with mean mu and variance sigma(mu)^2
- Suppose that an IQ test is given to an i.i.d. sample of students and that the score for each student in normally distributed with mean equal to their true IQ and variance sigma(e)^2 and is independent of the other student scores.
- Also assume that it is known that sigma(mu)^2/sigma(e)^2 = 9
- If the sample mean was 100, what is the best estimate for a student that scored 130?

The BLUP is:

Write Z = sigma(mu)^2/(sigma(mu)^2+sigma(e)^2)

Then the BLUP = Z*130 + (1-Z)*100

In other words, the BLUP is a credibility-weighted estimate of the student's score and the overall mean.

In this example, the BLUP = 127